Mol Gen Genomics (2005) 274: 428–441
DOI 10.1007/s00438-005-0037-0

ORIGINAL PAPER

**Young-Hoon Park · Magdy S. Alabady
Mauricio Ulloa · Brad Sickler · Thea A. Wilkins
John Yu · David M. Stelly · Russell J. Kohel
Osama M. El-Shihy · Roy G. Cantrell**

Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population

Received: 11 April 2005 / Accepted: 1 July 2005 / Published online: 27 September 2005

Abstract There is an immediate need for a high-density genetic map of cotton anchored with fiber genes to facilitate marker-assisted selection (MAS) for improved fiber traits. With this goal in mind, genetic mapping with a new set of microsatellite markers [comprising both simple (SSR) and complex (CSR) sequence repeat markers] was performed on 183 recombinant inbred lines (RILs) developed from the progeny of the interspecific cross *Gossypium hirsutum* L. cv. TM1 × *Gossypium barbadense* L. Pima 3-79. Microsatellite markers were developed using 1557 ESTs-containing SSRs (≥10 bp) and 5794 EST-containing CSRs (≥12 bp) obtained from ~14,000 consensus sequences derived from fiber ESTs generated from the cultivated diploid species *Gossypium arboreum* L. cv AKA8401. From a total of 1232 EST-derived SSR (MUSS) and CSR (MUCS) primer-pairs, 1019 (83%) successfully amplified PCR products from a survey panel of six *Gossypium* species; 202 (19.8%) were polymorphic between the *G. hirsutum* L. and *G. barbadense* L. parents of the interspecific mapping population. Among these polymorphic markers, only 86 (42.6%) showed significant sequence homology to annotated genes with known function. The chromosomal locations of 36 microsatellites were associated with 14 chromosomes and/or 13 chromosome arms of the cotton genome by hypoaneuploid deficiency analysis, enabling us to assign genetic linkage groups (LG) to specific chromosomes. The resulting genetic map consists of 193 loci, including 121 new fiber loci not previously mapped. These fiber loci were mapped to 19 chromosomes and 11 LG spanning 1277 cM, providing approximately 27% genome coverage. Preliminary quantitative trait loci analysis suggested that chromosomes 2, 3, 15, and 18 may harbor genes for traits related to fiber quality. These new PCR-based microsatellite markers derived from cotton fiber ESTs will facilitate the development of a high-resolution integrated genetic map of cotton for structural and functional study of fiber genes and MAS of genes that enhance fiber quality.

Keywords Cotton fiber transcriptome · Complex sequence repeat · Expressed sequence tag · *Gossypium* · Simple sequence repeat

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/10.1007/s00438-005-0037-0

Communicated by R. Hagemann

Y.-H. Park
Biotechnology Department, Sakata Seed America, Inc., Salinas, CA 93907, USA

M. S. Alabady · B. Sickler · T. A. Wilkins
Department of Plant Sciences, University of California, Davis, CA 95616, USA

J. Yu · R. J. Kohel
USDA-ARS, Germplasm Research Unit, College Station, TX 77845, USA

D. M. Stelly
Department of Soil and Crop Sciences, Texas AM University, College Station, TX 77843, USA

O. M. El-Shihy · M. S. Alabady
Plant Biotechnology Research Laboratory, Faculty of Agriculture, Cairo University, Cairo, Egypt

R. G. Cantrell
Cotton Incorporated, Cary, NC 27513, USA

M. Ulloa (✉)
WCIS Research Unit, Cotton Enhancement Program, USDA-ARS, 17053 N. Shafter Ave., Shafter, CA 93263, USA
E-mail: mulloa@pw.ars.usda.gov
Tel.: +1-661-7468009
Fax: +1-661-7461619

Introduction

Grown in more than 80 countries, cotton is a major crop and an important renewable resource, providing the world's leading natural fiber for the manufacture of textiles. Cotton belongs to the genus *Gossypium*, which consists of at least 45 diploid and five allotetraploid species (Percival et al. 1999; Ulloa et al. 2005a). *Gossypium hirsutum* L. (*Gh*, AD_1 genome) and *Gossypium barbadense* L. (*Gb*, AD_2 genome) are modern allotetraploid ($2n = 4x = 52$) cottons, which together represent the most extensively cultivated species worldwide. While *Gh* (upland cotton) is the most widely cultivated species—prized for its higher yield and wider environmental adaptation, *Gb* (Pima, Sea Island cotton, and Egyptian) boasts vastly superior fiber quality (Ulloa et al. 2005b). The allotetraploid species are thought to have formed about 1.1–1.9 million years ago (MYA) after a polyploidization event that brought together the genomes of diploids closely related to *Gossypium herbaceum* L. ($2n = 2x = 26$, A_1 genome) or *Gossypium arboreum* L. ($2n = 2x = 26$, A_2 genome), and *Gossypium raimondii* L. ($2n = 2x = 26$, D_5 genome), and were domesticated through extensive human selection (Wendel and Cronn 2003). A-genome diploid cottons produce spinnable fibers and have been cultivated, while D-genome species produce very short and appressed fibers. However, many quantitative trait loci (QTLs) for fiber traits have been associated with the D-subgenome in allotetraploid cottons (Paterson et al. 2003; Ulloa et al. 2005b), indicating that the D genome contributes to fiber morphogenesis and the determination of fiber properties.

The main goals of cotton breeding programs worldwide are the genetic enhancement of yield and, more recently, fiber quality. Improvement of fiber properties is required to keep pace with the rapid changes taking place in the technology of the manufacturing procedure. Key fiber quality traits, such as fiber elongation, length, fineness, and bundle strength, are controlled by QTLs, which complicate conventional breeding for fiber improvement (Jiang et al. 2000; Ulloa and Meredith 2000; Paterson et al. 2003; Mei et al. 2004; Ulloa et al. 2005b). Molecular markers offer efficient tools for dissecting QTLs affecting traits with complex genetic inheritance, and facilitate marker-assisted selection (MAS) and map-based cloning.

A high-density molecular map of fiber genes can assist breeding procedures by directly tagging the genes conferring traits of interest. To date, several genetic maps of cotton genomes have been constructed using diverse molecular markers and mapping populations (Reinisch et al. 1994; Ulloa et al. 2000; Ulloa et al. 2002; Lacape et al. 2003; Rong et al. 2004). Two interspecific genetic maps with wide genome coverage (>80%) derived from F_2 (Rong et al. 2004) and BC_1 progeny (Lacape et al. 2003) of *Gh* × *Gb* were recently reported, providing a framework for an integrated high-resolution map. However, a high-density molecular map composed of loci for fiber genes is still lacking in cotton, primarily due to the lack of fiber gene sequences, and the limited number of simple PCR-based DNA markers, such as microsatellites, available in the public domain.

Microsatellites or concatemeric repeats of short DNA sequences resulting from mutational effects of replication slippage are found in abundance and scattered throughout the genomes of eukaryotes (Tautz and Schotterer 1994). Generally, microsatellites are classified into two types: simple sequence repeats (SSRs) and complex sequence repeats (CSRs). A SSR is composed of one type of repeating unit, while a CSR contains stretches of two or more different repeat motifs, which can also vary in the number of repeating units, and therefore lead to an additional source of polymorphism. CSRs represent the second largest group of microsatellites in eukaryotic genomes, accounting for about 10% of microsatellites in the human genome (Weber 1990). Amplification of SSRs or CSRs by PCR reveals length polymorphisms of microsatellites resulting from differences in the number of repeat units. This simple PCR-based marker system is very valuable because of the co-dominant, highly variable and multiallelic nature of the markers it provides (Gupta et al. 1996). The occurrence of microsatellite loci in EST collections has been observed in *Arabidopsis* (Cardle et al. 2000) and several major crops (Chin 1996; Temnykh et al. 2000; Thiel et al. 2003) including tetraploid cotton (Saha et al. 2003; Qureshi et al. 2004). ESTs containing microsatellites offer a relatively simple and efficient means for the development of PCR-based microsatellite markers with functional gene information (Scott et al. 2000; Thiel et al. 2003).

A recently described set of ESTs from cotton fiber (Arpat et al. 2004) provides a valuable new resource for developing PCR-based DNA markers for fiber genes. The genetic complexity of the cotton fiber transcriptome is very high, comprising as much as 50% of the cotton genome (Arpat et al. 2004; Wilkins and Arpat 2005; Wilkins et al. 2005). This complexity translates to an estimated 36,000 or so homologous fiber genes in the At and Dt genomes of tetraploid species. The abundance and allelic diversity of microsatellites in coding sequences of cotton fiber genes has not been fully exploited as yet, although Han et al. (2004) recently published a genetic map based on 99 SSR markers (NAU) developed from fiber ESTs from *G. arboreum*. Thus, despite the economic importance of cotton fibers, only a limited number of fiber genes have been mapped, and there is a significant opportunity for further mining of the fiber EST database for markers to facilitate genetic mapping of fiber genes.

In the present study, new microsatellite markers were developed from a subset of 6000 SSR/CSR-containing fiber EST consensus sequences to construct the first fiber EST-microsatellite cotton genetic map from interspecific recombinant inbred lines (RILs). Assignment of fiber loci to chromosomes was accomplished using interspecific chromosomal substitution lines. As a preliminary assessment of fiber QTLs, the detection

of a putative QTL using F_2 phenotypic data is described. These new markers derived from cotton fiber ESTs will aid the development of a high-resolution integrated genetic map of cotton for structural and functional study of fiber genes and MAS for fiber-related genes.

Materials and methods

Identification of fiber ESTs containing microsatellites

The fiber EST database established for *G. arboreum* L. cv AKA8401, which comprises ~14,000 consensus sequences, was searched for simple (SSR) and complex (CSR) dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats using a modified version of Sputnik (Abajian, Washington University; http://abajian.com/sputnik) compiled with a recursive algorithm: minimum unit length = 2, maximum unit length = 5, exact match points = 1, error match points = −6, match minimum score = 8, match fail score = −1, and maximum recursion = 5. Consensus sequences containing SSRs (≥10 bp) and CSRs (≥12 bp) were annotated with the most significant BlastX hit against GenBank's non-redundant (nr) database using custom Perl scripts and PERL5 interface modules, and the results have been posted on the web (http://cfgc.ucdavis.edu/ssr).

Development of EST-derived microsatellite markers

From a total of 7351 microsatellite-containing fiber ESTs, a subset of 1761 fiber genes were selected for PCR primer design. The selection mainly targeted ESTs, which were annotated by homologous gene functions. ESTs that had previously been mapped on the basis of RFLPs in cDNAs (Rong et al. 2004) were also selected in order to convert them into PCR-based markers and compare their allelism. PCR primer-pairs flanking SSR/CSRs were designed using PRIMER3 software v.0.2c based on the following core parameters: primer length = 18–25 bp, size of amplification product = 150–300 bp, minimum GC content = 40%, GC clamp = 2, optimum melting temperature = 60°C. The range of the parameters was modified when suitable primer pairs were not initially retrieved by the software. Each primer-pair was numbered and assigned the designation MUSS and MUCS for SSRs and CSRs, respectively. PCR primers were synthesized and desalted for purification by Proligo (Boulder, Colo.).

PCR amplification was performed in a total volume of 20 µl containing 20 ng of genomic DNA, each primer (forward and reverse) at 0.1 µM, 1×PCR buffer, 0.2 mM dNTPs, and 1 U of Taq polymerase (Amplitaq, Applied Biosystem, Foster City, CA, USA) with the following cycling profile: 1 cycle of 2 min at 94°C, 10 cycles of 15 s at 94°C, 30 s at 60°C (decreasing in steps of 0.5°C/cycle for cycles 2–10), and 1 min at 72°C, followed by 35 cycles of 15 s at 94°C, 30 s at 55°C, and 1 min at 72°C. PCR products were separated on a 3% Super Fine Resolution (SFR) agarose (Amresco, Solon, OH, USA) gel containing TBE at 80 V for 4–5 h, and visualized with Alphaimager (Alpha Innotech, San Leandro, CA, USA) software v. 5.5 after staining with ethidium bromide. Primer-pairs resulting in discrete PCR banding patterns denoting a molecular marker were scored.

Plant materials used for polymorphism evaluation

The successful amplification of PCR products by EST-derived microsatellite primer-pairs, and initial screening for polymorphism, were conducted using a DNA panel composed of eight different cultivars and accessions of *Gossypium* spp. (Table 1). The genotypes in the survey panel included four cultivars of *Gh* and *Gb*, which were used as parents for two intraspecific mapping populations, and one accession for each of four A-genome and D-genome diploid species. Polymorphic microsatellites scored between *Gh* and *Gb* were selected, and subsequently tested for marker polymorphism information content (PIC) using a second DNA panel consisting of 12 cultivars and accessions of *Gh* and *Gb*, one allotetraploid interspecific inbred line, and two accessions of *Gossypium tomentosum* Nuttall ex Seemann and *Gossypium mustelinum* Miers ex Watt. The PIC value refers to the value of the marker for detecting polymorphism based on the number of alleles and their distribution frequency within a population (Botstein et al. 1980). The PIC value for the newly developed markers between *Gh* and *Gb* was estimated using the equation of Anderson et al. (1993):

$$\mathrm{PIC} = 1 - \sum_{j=1}^{n} P_{ij}^2$$

where P_{ij} is the frequency of the jth microsatellite pattern for marker i and the summation covers n patterns. Genomic DNA was extracted following the method described by Ulloa and Meredith (2000).

Chromosomal assignment of microsatellite markers

Hypoaneuploid stocks developed from an interspecific cross between *Gh* cv. Texas Marker-1 (TM1) and *Gb* cv. Pima 3-79 (3-79) were obtained from the Cotton Cytogenetics Collection at Texas A&M University, and used for chromosomal assignment of SSR markers (Stelly et al. 1993; Liu et al. 2000; Ulloa et al. 2005b). The cytogenetic stocks consisted of 15 primary monosomic, 25 monotelodisomic, and six tertiary monosomic F_1 substitution lines in which an entire chromosome, all or most of a chromosome arm, or specific chromosome segments of *Gh* are missing, respectively, while the intact homologous chromosome of *Gb* is present. Selection of microsatellite markers for PCR amplification from

Table 1 List of cotton cultivars and accessions used to study microsatellite marker variation in fiber ESTs

Cultivar/Accession[a] Sample panel 1	Taxonomy	Genome
MD51ne-Okra	*G. hirsutum* L.	AD_1
Fiber Max 832	*G. hirsutum* L.	AD_1
Pima S-6	*G. barbadense* L.	AD_2
Pima 89590	*G. barbadense* L.	AD_2
A_{1-5}	*G. herbaceum* L.	A_1
A_{2-8}	*G. arboreum* L.	A_2
D_{1-1}	*G. thurberi* Todaro	D_1
D_{2-2}	*G. harknessii* Brandegee	D_2
Sample panel 2		
03MU2010	*G. hirsutum* L.	AD_1
9456-0	*G. hirsutum* L.	AD_1
Acala Maxxa	*G. hirsutum* L.	AD_1
MD51ne-Normal	*G. hirsutum* L.	AD_1
Phytogen 72	*G. hirsutum* L.	AD_1
Shafter-SC01	*G. hirsutum* L.	AD_1
SL-1-7-1	*G. hirsutum* L.	AD_1
Tamcot HQ95	*G. hirsutum* L.	AD_1
TM1	*G. hirsutum* L.	AD_1
03MUGH3	*G. barbadense* L.	AD_2
Pima 3-79	*G. barbadense* L.	AD_2
Pima S-7	*G. barbadense* L.	AD_2
NM24016	Interspecific inbred (*Gh* × *Gb*)	
WT936	*G. tomentosum* Nuttall ex Seemann	AD_3
W400	*G. mustelinum* Miers ex Watt	AD_4

[a]The PIC values were calculated using the information on allelic variation obtained from DNAs from both sample panels
Sample panel 1 was used to test PCR amplification and polymorphisms for 1232 primer pairs. The primer pairs that detected polymorphism between *G. hirsutum* L. (*Gh*) and *G. barbadense* L. (*Gb*) in sample panel 1 were tested further using DNAs from sample panel 2

DNA of aneuploid stocks was limited to polymorphic markers that yielded PCR products of different sizes from TM1 and 3-79. Microsatellite markers were assigned to the deficient chromosome of the aneuploid stock from which the TM1 allele was not amplified. Microsatellite loci that localized to one of the chromosomes (Chr.) 1–13 were assigned to the A-subgenome (At), whereas loci localized to chromosomes 14–26 were assigned to the D-subgenome (Dt).

Construction of genetic linakge map

Genomic DNA was prepared from 183 RILs generated from a cross between an Upland cotton (*Gh* TM1) and a Pima cultivar, which yields fibers of superior quality (*Gb* 3-79). The development of the original F_2 population is described by Kohel et al. (2001). For genotyping of the 183 RILs, a total of 95 SSR and 72 CSR polymorphic markers identified between TM1 and 3-79 were selected from this study, as well as 88 publicly available SSR markers (73 BNL, 6 CIR, and 9 JESPR) spanning the cotton genome as framework markers (Nguyen et al. 2004; Han et al. 2004), and used as anchored markers for chromosome assignment of our linkage groups (LG). PCR and gel electrophoresis were conducted as described above. The genotypes of TM1, 3-79 and heterozygotes were scored as A, B, and H, respectively. Multiple or duplicated loci detected by a single primer-pair were designated by adding lower case letters after the name of the primer-pair.

A genetic linkage map was constructed using JoinMap 3.0 (Stam and Van Ooijen 1995). The Kosambi map function was used to convert recombination frequency to genetic map distance (centiMorgan, cM). Most of the LG were determined at LOD scores ≥6. Deviation from the 1:1 segregation ratio expected for RILs was assessed for each marker using χ^2 statistics.

The mapped markers were categorized based on their homologous gene functions. The gene function information of the mapped markers was obtained from the Gene Ontology (GO) web site (www.geneontology.org). A Perl script has been written which allows the sequences of the mapped markers to be compared with the UniProt protein database (www.pir.uniprot.org). The GO numbers for the best homologous hits were used to find molecular function, cellular component and biological process ontology for these markers.

QTL analysis of fiber quality-related traits

In a previous study, the detection of 13 QTLs for bundle fiber strength (*Sf*), length at 2.5% (2.5% *Lf*), and fineness (*Ff*) have been reported using RFLP and RAPD genotypes and phenotypic data obtained from the original F_2 progeny from which the RILs used in this study were derived (Kohel et al. 2001). Since the phenotypic data from RILs were not available at the time this study was completed, the QTL analysis reported here was conducted using the same phenotypic data from the F_2 progeny in an attempt to validate the chromosomal location of previously reported QTLs by using RILs and microsatellite markers. In addition to the three fiber quality traits studied by Kohel et al. (2001), two other traits, fiber elongation (*Ef*) and fiber length at 50% (50% *Lf*) were included in this study. *Ef* is an estimate of the elasticity of a bundle of fibers. The 50% *Lf* is the distance spanned by 50% of fibers scanned with the fibrograph, while other measurements of fiber quality traits were described previously in Kohel et al. (2001).

QTL analysis was conducted using the programs MapQTL 4.0 (Van Ooijen and Maliepaard 1996) and WinQTLCart 2.0 (Wang et al. 2004). Simple marker analysis (SMA) fitting of data to the simple linear regression model was performed to estimate the *F* statistics of each marker for association to a QTL. Composite interval mapping (CIM) (Zeng 1994) was performed in a stepwise forward regression procedure to detect putative QTL positions using a likelihood ratio (LR) test statistic. Empirical threshold values for LR were determined after 1000 permutation tests for all traits (Churchill and Doerge 1994).

Results

Abundance, distribution, and gene homologues of microsatellites derived from fiber EST sequences

From the *G. arboreum* L. fiber EST database of ~14,000 consensus sequences, a total of 1557 SSRs, and 5794 CSRs with perfect or imperfect repeats interrupted by non-repetitive bases, were identified and classified based on their motifs and number of repeats. Among the four repeat types identified in this search (dinucleotides, trinucleotides, tetranucleotides, and pentanucleotides), trinucleotides occurred most often (48.6% for SSRs and 52.0% for CSRs), while dinucleotides, tetranucleotides, and pentanucleotide repeats were found at lower frequencies. The most common dinucleotide motif was AT (TA), while TTC (AAG) was the most frequent trimeric repeat unit. For tetranucleotides and pentanucleotide repeat motifs, AAAN (TTTN), and AAAAN (TTTTN), respectively, was most frequent. Repeat numbers varied most for dinucleotide motifs, ranging from 5 to 45, with an average repeat number of 8.3. In total, 1066 SSR-containing ESTs showed significant homology (E-value $< 1\times10^{-20}$; BLASTX) to putative genes with known or unknown function, whereas 491 sequences resulted in no significant hits when compared with GenBank's nr database. The cDNA clones of 84 SSR-containing fiber ESTs were mapped previously as RFLPs (Rong et al. 2004).

The motif structure of CSRs was highly diverse and difficult to classify as a result, although a few general trends were observed. The most frequent di-, tri-, tetra-, and penta-repeats upstream motifs were $(CT/TC)_n$, $(CTT/TCT/TTC)_n$, $CTTT/TCTT/TTTC)_n$, and (ATTT/TATTT/TTATT/TTTTA), while downstream motifs were $(AT/TA)_n$, $(CTT/TCT/TTC)_n$, $(AAAG/AAGA/AGAA/GAAA)_n$, and $(AAAAG/AAGAA/GAAAA)_n$, respectively. CSRs of 18 bp in length were the most abundant (183 sequences), while CSRs longer than 33 bp were found in only three ESTs. From 1761 EST-derived CSRs, 1566 were annotated with the most significant BLASTX hit, with 241 matching cDNA clones previously mapped as RFLPs (Rong et al. 2004).

Evaluation of marker polymorphisms

SSR-containing (761) and CSR-containing (1000) cotton fiber EST sequences were selected for primer design. The SSR-containing EST sequences included all ESTs with significant BLASTX hits, as well as all ESTs for which cDNA clones had previously been mapped (Rong et al. 2004). A subset (63.8%) of the CSR-containing ESTs with significant BLASTX hits, and all ESTs with cDNA-RFLP map positions, were selected. Primer-pairs flanking SSRs or CSRs were successfully designed for 1232 unique ESTs (Table 2). Primer design was not possible for 529 ESTs, because the microsatellite was located at the border of the sequenced DNA fragment, or surrounded by AT-rich repetitive DNA. PCR amplification was successful with more than 80% of the primer-pairs tested, and markers obtained with 1019 primer-pairs were easily scored across a panel of eight genotypes containing diploid and tetraploid accessions and cultivars (Table 2), and these primers were included in our analyses. The majority of these primer-pairs amplified a DNA fragment of the size expected on the basis of *G. arboreum* cDNA sequences. However, 290 primer-pairs produced PCR amplicons that were larger than expected, while 17 primer pairs produced shorter amplicons. Judging from homologous *Arabidopsis* gene models, the variation in size of amplicons is due to the presence of introns in genomic DNA and/or to a varying number of repeats, as shown in Han et al. (2004).

A total of 1522 DNA PCR fragments were amplified from six species using 1019 primer-pairs (Fig. 1a, Table 2). The number of polymorphic fragments amplified across six cotton species ranged from two to nine, with an average number of 2.3 PCR products. As expected, the majority of the primer-pairs amplified a single fragment from the diploids. In allotetraploids, 25% of the primer-pairs amplified ≥2 fragments, probably derived from duplicated and/or homologous loci in the two subgenomes. The numbers of markers that were polymorphic within or between the species are summarized in Table 2. In all, 168 of the primer-pairs amplified single (monomorphic) fragments of uniform size from all species.

Of the 562 polymorphic markers identified, only 202 (35.9%) were polymorphic between *Gh* and *Gb*, and thus suitable for interspecific mapping. Only 86 (42.6%) of these markers were significantly homologous to genes of known function, especially to genes associated with *Ef* (Ji et al. 2003; Arpat et al. 2004). Polymorphism was low within each of the two allotetraploid species (Table 2). Among the different types of SSRs, the highest rate of polymorphism (73.1%) was observed for markers containing dinucleotide repeats among the six species. For trinucleotide repeats, the rate of polymorphism across all accessions was lowest (42.3%). The relationships between types of motif, number of repeats and polymorphism level are shown in Fig. 2. A general trend was observed which is characterized by an increase in the level of polymorphism as the number of repeats increases for all types of motifs across all six species. The correlation between the level of polymorphism and number of repeats, however, was not in evidence for dinucleotides and trinucleotides between *Gh* and *Gb*, although a polymorphism rate of 41% was found for trinucleotides when the number of repeats was greater than seven (data not shown).

PIC values were assessed for 173 MUSS/MUCS polymorphic primer-pairs between *Gh* and *Gb* using 23 DNA samples in order to assess marker diversity and judge the utility of these markers for fingerprinting

Table 2 Summary of the development of PCR-based simple sequence repeat (SSR) and complex sequence repeat (CSR) markers from the cotton (*G. arboreum*) fiber EST database and the evaluation of polymorphism

Parameter	Marker type		Total
	SSR	CSR	
Total number of microsatellite-containing ESTs listed	1557	5794	7351
Primer design			
ESTs screened for primer design	761	1000	1761
Primer pairs designed	608	624	1232
PCR amplification			
Successful primer pairs	496	523	1019
Product of expected size[a]	344	368	712
Product of unexpected size	152	155	307
Unsuccessful primer pairs	112	101	213
Polymorphic primer pairs[b]			
Across six species	311	251	562
Between *Gh* and *Gb*	114	88	202
Within *G. hirsutum*	7	7	14
Within *G. barbadense*	4	11	15
Between A and D-genome diploids	106	239	345
Within A-genome diploids	81	38	119
Within D-genome diploids	120	85	205

[a]PCR fragments of approximately the sizes expected from *G. arboreum* EST sequences
[b]Cotton accessions of each species for DNA samples were described in materials and methods

and genetic mapping (Table 1). PIC values ranged from 0.73 to 0.12, with an average of 0.42. MUCS258 showed the highest PIC value (0.73) and the lowest value (0.12) was found for 11 markers (Supplementary Tables S1 and S2). The markers with higher PIC values have greater potential to reveal allelic variations, and can be preferentially used for choosing appropriate parental lines for molecular mapping or assessing genetic diversity within a population (Anderson et al. 1993).

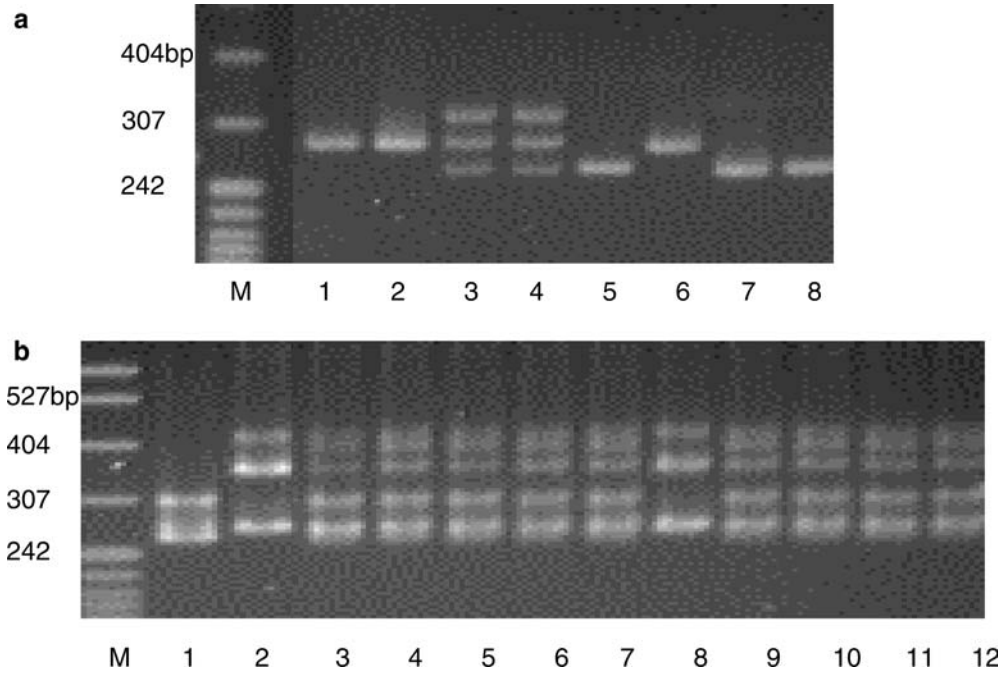

Fig. 1 Agarose gel (SFR) image showing PCR products amplified by two cotton fiber EST-derived microsatellite markers, MUSS210 (**a**) and MUSS599 (**b**) (Table 4). **a** Marker evaluation using a DNA sample panel comprising eight cultivars and accessions (Table 1) of *Gossypium*. *M* size standard marker. Lanes: *1* MD51ne (*G. hirsutum* L., AD_1), *2* Fiber Max 832 (*G. hirsutum* L., AD_1), *3* Pima S-6 (*G. barbadense* L., AD_2), *4* Pima 89590 (*G. barbadense* L., AD_2), *5* A_{1-5} (*G. herbaceum* L., A_1), *6* A_{2-8} (*G. arboreum* L., A_2), *7* D_{1-1} (*Gossypium thurberi* Todaro, D_1), *8* D_{2-2} (*Gossypium harknessii* Brandegree, D_2). **b** Chromosomal assignment of microsatellite marker using monosomic stocks. *M* size standard marker. Lanes: *1 G. hirsutum* L. cv. Texas Marker 1 (TM1), *2 G. barbadense* L. cv. Pima 3-79 (3-79), *3–12* monosomic lines deficient for one chromosome or chromosome segment. The TM1 allele of 307 bp (*arrow*) is missing from the monosomic line (genotype: HO2) that is deficient for chromosome 2 (lane 8), indicating that the locus is located on chromosome 2

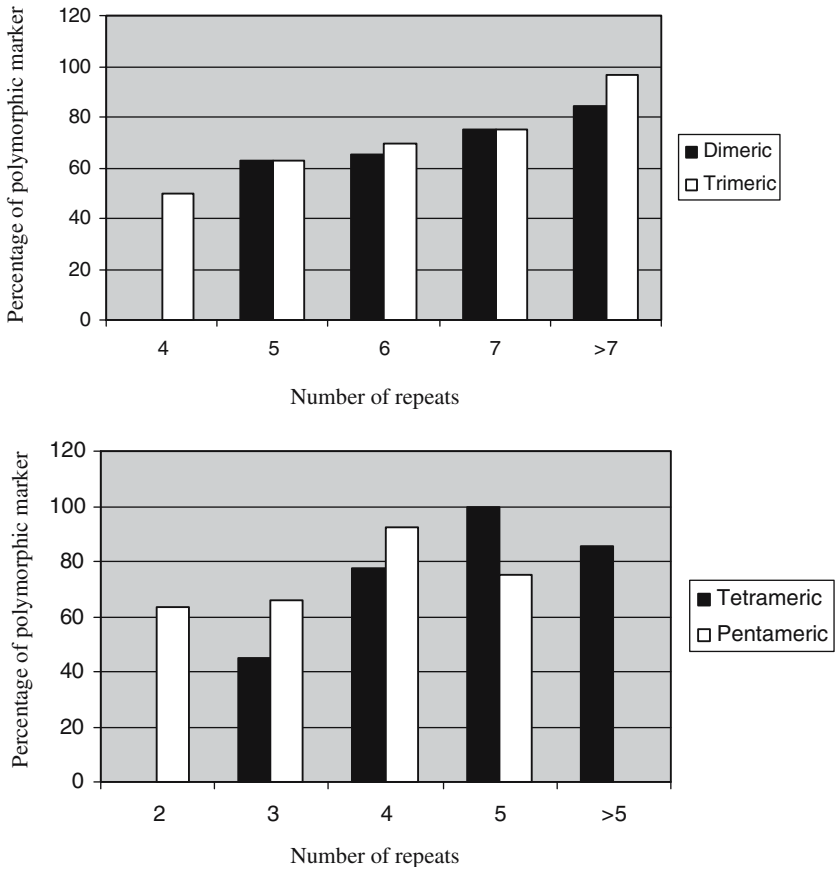

Fig. 2 Distribution of the types of repeat among 496 fiber EST-derived SSR markers, and relationships between motif length, number of repeats, and level of polymorphism among eight cultivars and species of cotton (*Gossypium*) (Table 1)

Localization of microsatellite DNA fragments to individual chromosomes

A total of 101 (58 SSR and 43 CSR) EST markers that were polymorphic between *Gh* and *Gb* were used to amplify DNA fragments from 46 F_1 substitution aneuploid lines and two euploid parents, TM1 and 3-79, to assign chromosomal locations to the markers (Stelly 1993; Liu et al. 2000; Ulloa et al. 2005b). Thirty-six microsatellite markers failed to detect TM1 fragments in the aneuploid stocks (Fig. 1b). These markers could be assigned to 14 chromosomes (Chr. 01, 02, 03, 07, 09, 10, 11, 12, 15, 16, 17, 18, 20, and 26) and 13 chromosome arms (Chr. 01*sh*, 03*Lo*, 03*sh*, 07*Lo*, 09*Lo* 10*sh*, 10*Lo* 11*sh*, 15*sh*, 16*Lo*, 18*Lo*, 20*sh*, and 20*Lo*) (Supplementary Tables S1 and S2). Seven SSR (MUSS021, 068, 124, 181, 280, 298, and 343) and three CSR (MUCS160, 267, and 405) markers amplified DNA fragments from both TM1 and 3-79 parents and from all monosomic stocks, and therefore could not be assigned to chromosomes and are likely to be located on chromosomes for which deficient lines were not available. Sixteen SSR and eight CSR markers were assigned to eight chromosomes or chromosome arms of the A subgenome (Chr. 01, 02, 03, 07, 09, 10, 11, and 12), and seven SSR and five CSR markers were assigned to six chromosomes or chromosome arms of the D subgenome (Chr. 15, 16, 17, 18, 20, and 26) (Supplementary Tables S1 and S2).

Genetic map construction

A total of 183 RILs were genotyped using 253 PCR primer-pairs (91 MUSS, 62 MUCS, 85 BNL, 6 CIR, and 9 JESPR). A total of 209 primer-pairs (77 MUSS, 44 MUCS, 73 BNL, 6 CIR, and 9 JESPR) generated discrete banding patterns for reliable genotyping of the RILs, while the rest of the primer-pairs were not polymorphic in this mapping population, and/or genotyping could not be completed due to weak PCR amplification or unsuccessful gel separation of alleles of similar size in this gel system. A total of 224 loci were scored and used for the development of a genetic linkage map, and 31 markers were not linked to any linkage group. Significant departures from the expected 1:1 segregation ratio were detected for three loci at $P \leq 0.05$, seven at $P \leq 0.01$, and 28 at $P \leq 0.001$. Among the loci that displayed a pronounced degree of segregation distortion ($P \leq 0.001$), 23 showed skew toward the allele from the maternal parent, TM1.

A genetic map composed of 40 LG with 193 loci (73 MUSS, 48 MUCS, 59 BNL, 8 JESPR, and 5 CIR markers) was constructed (Fig. 3). The map spanned 1277 cM (689 and 490 cM for the At and Dt genomes, respectively), covering approximately 27% of the allotetraploid cotton genome (2250 Mb; Arumuganathan and Earle 1991) with an average distance of 6.62 cM between markers. BNL and MUSS/MUCS markers

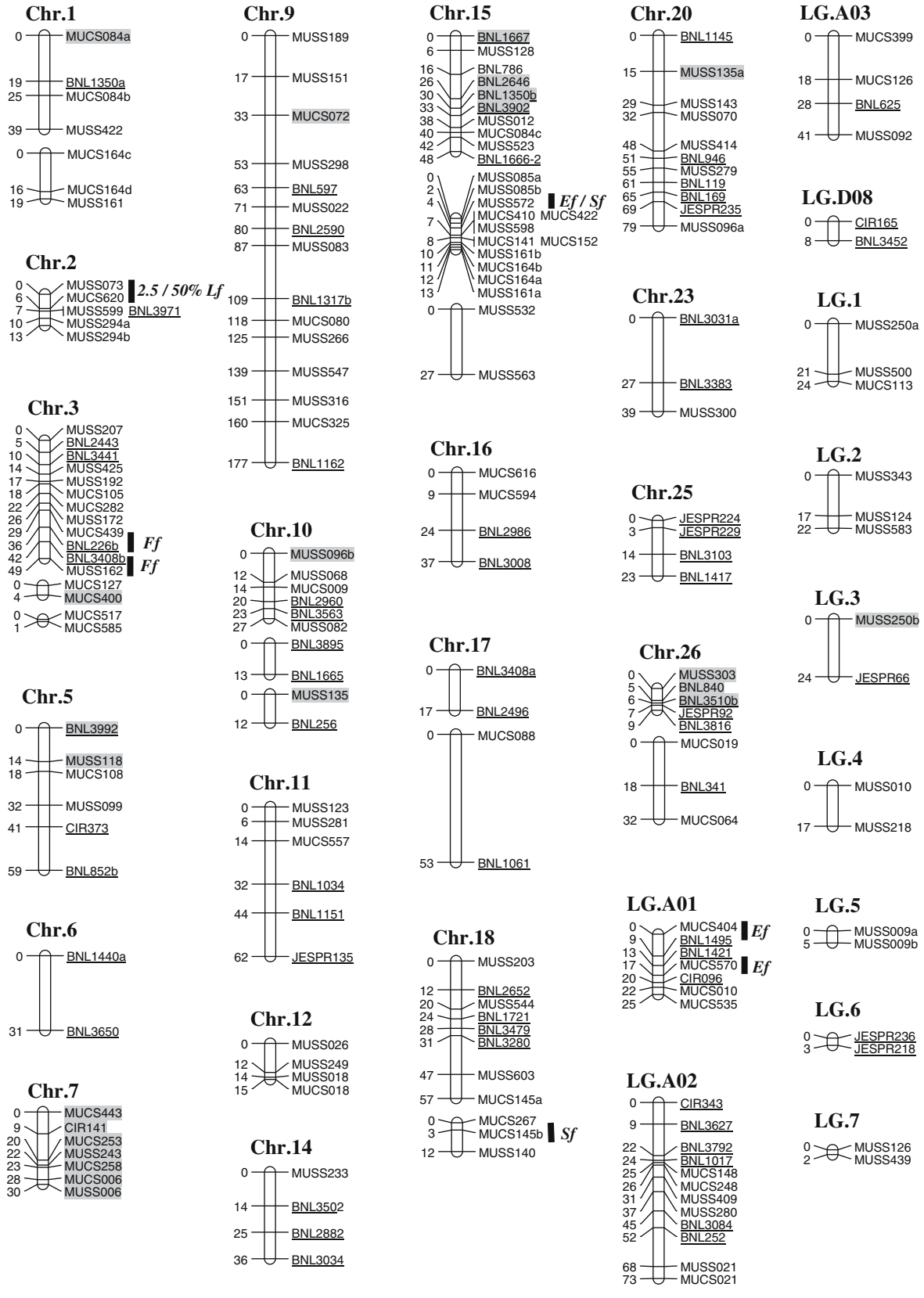

Fig. 3 Genetic linkage map of cotton constructed from 183 recombinant inbred lines (RILs) derived from the interspecific cross *Gh cv.* TM1 × *Gb cv.* Pima 3-79 using 193 fiber EST-derived microsatellites and framework SSR markers (BNL, JESPR, and CIR) (*underlined*) to anchor the map. The markers that show significant segregation distortion ($x^2 > 10$) are indicated by *gray blocks*. Map intervals for putative QTLs for five fiber quality-related traits [fiber enlongation (*Ef*), fiber bundle strength (*Sf*), fiber length at 2.5 and 50% (2.5 and 50% *Lf*), and fiber fineness (*Ff*)] are represented by *solid vertical bars*

Table 3 Fiber quality-related QTL detected by composite interval mapping (CIM) in cotton

Trait	QTL	Program[a]	Map location (subgenome)	Nearest marker	LOD	Effect (%)[b]
Fiber elongation	*Ef*	Q	LG.A01	MUCS404	2.1	5.5
		Q/M	LG.A01	MUCS570	2.1	5.4
		M	Chr.15 (D)	MUSS572	2.1	5.7
Fiber bundle strength	*Sf*	Q/M	Chr.15 (D)	MUSS572	2.7	5.4
		Q	Chr.18 (D)	MUCS145b	2.6	6.3
Fiber length at 2.5%	2.5% *Lf*	Q	Chr.02 (A)	MUCS620	2.0	5.2
Fiber length at 50%	50% *Lf*	Q/M	Chr.02 (A)	MUCS620	2.9	6.6
Fiber fineness	*Ff*	Q	Chr.03 (A)	BNL226b	2.6	6.2
		Q/M	Chr.03 (A)	BNL3408b	2.7	7.9

For QTLs detected by both programs (Q/M), *Q* WinQTLcart 2.0, *M* MapQTL 4.0. the LOD scores are based on the MapQTL data
[a]QTL mapping program(s) with which the QTLs were detected
[b]Percentage of phenotypic variance explained; effects of individual QTLs are additive

anchored to the same chromosome were mapped to the same linkage group, demonstrating the reliability of the map (Fig. 3). Fourteen linkage groups containing 100 loci were assigned to ten chromosomes of At, and the other 14 linkage groups containing 75 loci were assigned to nine chromosomes of the Dt genome (Supplementary Table S1 and Fig. 3).

Loci showing significant segregation distortion ($P \leq 0.001$) clustered to several chromosomal regions, with the exception of four loci that were not grouped (Fig. 3). It is interesting to note that allelic segregation of all seven loci mapped to Chr. 07 was skewed towards the TM1 allele (Fig. 3). Multiple PCR fragments amplified by a single primer-pair (MUCS164c-d, MUSS294a-b, MUSS085a-b, MUSS161a-b, MUSS164a-b, and MUSS009a-b) mapped to duplicated loci within a chromosome (Fig. 3). Duplicated loci at homologous At and Dt chromosomes were also evident for several markers; MUSS096a-b (Chr. 10 and 20), MUCS084b-c (Chr. 01 and 15), and MUCS164ab-cd (Chr. 01 and 15) primer-pairs (Fig. 3). Two BNL marker loci, BNL1034 and 1151 have been reported on LG.A03 (Lacape et al. 2003) and this linkage group had not been assigned to a chromosome. In our map, these loci were tightly linked (LOD > 6.0) to MUSS123 and JESPR135, which were assigned to Chr. 11 by hypoaneuploid stock screening and previous genetic mapping, respectively, indicating a possible homologous relationship between LG.A03 and Chr. 11.

Categorization of gene functions

Classification of mapped markers based on gene functions revealed that the majority of the mapped MUSS and MUCS markers were associated with genes belonging to two major GO categories (molecular function and/or biological process) (Supplementary Table S1 and Fig. 4). Of 92 such markers, four assigned to the category "cellular component" (subcellular structures, locations, and macromolecular complexes) are mainly associated with nuclear and viral envelopes. Protein transport and folding, cell adhesion, transcriptional regulation and steroid metabolism are represented by six markers categorized under "biological process." The markers in the "molecular function" category primarily belong to metabolism-related genes, including the seven novel markers for kinase, hydrolase, catalytic enzyme, oxidoreductase, and endopeptidase inhibitor activities, in addition to metal ion binding.

Analysis of QTLs for fiber quality

The phenotypic values for *Ef* and fiber length at 50% (50% *Lf*) were analyzed using WinQTLcart. All the values showed a normal distribution among 183 F_2 progeny, as is typical for traits controlled by QTLs (data not shown). As expected for extra long staple *Gb* cotton,

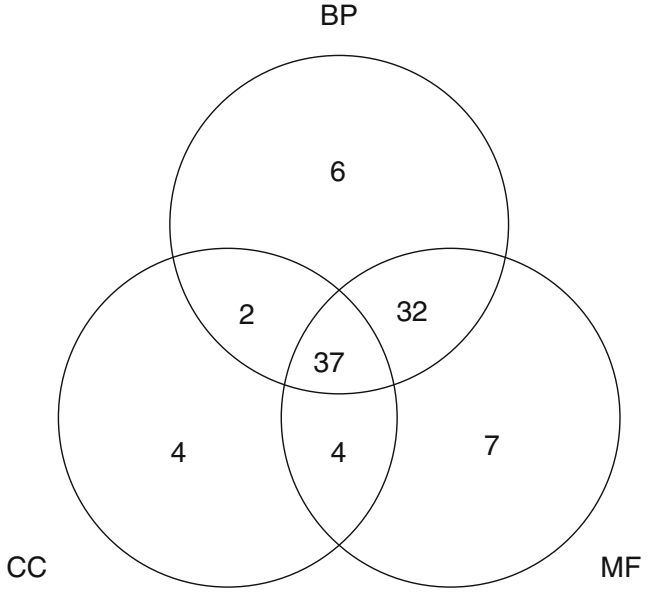

Fig. 4 Venn diagram showing the associated Gene Ontology categories of the mapped cotton fiber EST-derived SSR/CSR markers. *BP* biological process, *CC* cellular component, *MF* molecular function

Table 4 Fiber EST-derived microsatellites and cDNA-RFLP DNA markers developed from *G. arboreum* EST sequences and the comparisons of their locations on the three different genetic maps

EST sequence accession No.[a]	EST-SSR/CSR map			cDNA-RFLP map (Rong et al. 2004)		EST-SSR map (Han et al. 2004)	
	Loci[b]	Hypoaneuploid[c]	Linkage map[d]	Loci[b,e]	Linkage map[d]	Loci	Linkage map
CON_001_05453	MUSS006	Chr. 07*Lo*	Chr. 07			NAU1362	Chr. 07
CON_001_05991	MUSS009a-b		LG.05-05			NAU1236	LG.05
CON_001_06289	MUSS011		Ungrouped			NAU933	Ungrouped
CON_001_13096	MUSS073	Chr. 02	Chr. 2			NAU1072	Chr. 02
CON_002_00555	MUSS096a-b		Chr. 20-10			NAU1280	Chr. 20
CON_002_02244	MUSS118		Chr. 5			NAU861	Chr. 05
CON_005_03670	MUSS140	Chr. 18*Lo*	Chr. 18	Gate1CD12aE3C	Chr. 18	NAU1215	LG.A01
CON_008_04099	MUSS143	Chr. 10*sh*	Chr. 20			NAU904	Chr. 20
CON_001_07656	MUSS161a-b-c	Chr. 01*sh*	Chr. 15-15-01	Gafb14K15a-bE5C	Chr. 04-15		
CON_001_10145	MUSS172		Chr. 03			NAU1190	Chr. 03
CON_003_03068	MUSS218		LG.04	Gate3BF08a-bE6C	LG.A01-Chr. 18		
CON_011_04291	MUSS243	Chr. 07*Lo*	Chr. 07	Gate4BG08a-bE6D	LG.08-Chr. 16		
CON_014_04369	MUSS250a-b		LG.01-03	Unig27C04E6C	Chr. 02		
CON_043_04504	MUSS266		Chr. 09	Gate1AG08	LG.A01		
CON_001_04799	MUSS298		Chr. 09			NAU864	Chr. 09
CON_008_04060	MUCS145a-b		Chr. 18-Chr. 18	Gate4CG12E6D	Chr. 18		
CON_001_07656	MUCS164a-b-c-d		Chr. 15-15-01-01	Gafb14K15a-bE5C	Chr. 04-15		
CON_010_04251	MUCS253		Chr. 07	Gate4CG05a-bE3R	LG.D07-Chr. 04		
CON_011_04291	MUCS258	Chr. 07*Lo*	Chr. 07	Gate4BG08a-bE6D	LG.A05-Chr. 16		
CON_025_04463	MUCS282	Chr. 03*Lo*	Chr. 03	Gate4BE06E5D	Chr. 15		

[a]Consensus EST sequences listed at www.http://cfgc.ucdavis.edu
[b]Duplicated loci detected by the marker are indicated by *lower case letters*, which refer to the corresponding chromosomal locations were linked by *hyphens*
[c]Chromosomal location assigned by screening hypoaneuploid stock
[d]The map locations of duplicated loci are shown in the same order as hyphened in the second column
[e]Only 11 markers out of 29 compared for the map locations were represented as examples

3-79 showed better fiber quality, lower *Ef* (6.7%) and higher 50% *Lf* (35.6 mm) compared to TM1 (7.0% for *Ef* and 28.9 mm for *Lf*), respectively. Mean values for *Ef* and 50% *Lf* from F_2 progeny were 9.0 and 30.0 mm, respectively. Phenotypic analysis of the other three traits, *Sf*, fiber length at 2.5% (2.5% *Lf*) and *Ff* was described by Kohel et al. (2001).

Preliminary QTL analysis for fiber-related traits revealed one QTL on Chr. 02 for fiber length (50% SL) at a significant LOD level (2.9), which was higher than the empirical threshold (2.8) determined by 1000 permutation tests at $P < 0.05$ (Table 3). The MUCS620 marker linked to this SL fiber QTL is homologous to an endo-β-1,4-glucanase (*E*-value = 0.0), an enzyme that plays an important role in cell extension during rapid polar elongation of developing fibers (Arpat et al. 2004). For other fiber traits, permutation tests gave LOD scores of 2.8 (for *Ef* and 2.5% *Lf*) and 2.9 (for *Sf* and *Ff*) as the significant thresholds, although no QTL was detected above these LOD scores. At LOD≥2.0, however, a total of eight potential QTLs for *Ef*, *Sf*, 2.5% *Lf* and *Ff* were detected on Chr. 02, 03, 15 and 18, and LG.A01 (Fig. 3 and Table 3). From single marker analysis using MapQTL 4.0 and WinQTLCart 2.0, all of these loci were associated with the fiber traits at $P < 0.01$. QTLs detected on the chromosomes mentioned above were reported in previous studies for *Ef*, *Sf*, and *Ff* (Kohel et al. 2001; Paterson et al. 2003; Ulloa et al. 2005b).

Comparisons of map locations for homologous fiber EST-derived DNA markers

Redundancy for these new markers was evaluated by identifying previously reported NAU SSR (Han et al. 2004) and cDNA-RFLP markers (Rong et al. 2004) derived from the same fiber ESTs. These homologous markers also provide an opportunity to compare the chromosomal locations in relation to genetic maps. The EST sequences corresponding to the 99 fiber EST-derived SSR markers reported by Han et al. (2004) were aligned against consensus EST sequences of MUSS markers using ClustalW (www.http://services.bioasp.nl/blast/cgi-bin/clustal.cgi.). SSR markers showing EST identity were earmarked, and their corresponding locations on the genetic maps compared. The potential overlap between fiber genes on the RFLP map (Rong et al. 2004) and our fiber EST-derived SSR map positions was also examined.

A total of 38 MUSS/MUCS markers showed EST sequence homology to either NAU or cDNA-RFLP markers (Table 4). For these homologous markers, the chromosomal locations reported in the three different genetic maps were compared. Among ten homologous MUSS-NAU marker pairs detected in this survey, seven pairs mapped to the same chromosomes, while one homologous marker pair, MUSS140-NAU1215, mapping to different locations on Chr. 18 and LG.A01,

respectively (Table 4). However, for two marker pairs, MUSS009–NAU1236, and MUSS011–NAU933, comparison of chromosomal location was not possible since the markers were not linked to any group or mapped at unknown linkage groups (Table 4).

A search of the cDNA-RFLP map (Rong et al. 2004) identified 29 MUSS/MUCS markers that were also mapped as cDNA clones (examples are listed in Table 4). Nine MUSS/MUCS-RFLP homologous marker pairs mapped to the same or homologous chromosomes, while 20 pairs (18% of the loci in our map) mapped to different locations. The reason for this discrepancy is unknown. However, RFLP and microsatellites are very different marker systems, and it is possible that more than one copy of the cDNA sequence used is represented in the genome as multigene families, pseudogenes, or silent genes, while the SSR/CSR-markers target specific DNA sequences and motifs.

Discussion

Improving cotton fiber quality while maintaining fiber yield is a challenging task for conventional breeding, because of the negative association between these two traits and the genetic complexity of fiber QTLs (Paterson et al. 2003; Ulloa et al. 2005b; Shen et al. 2005). The genetic enhancement of fiber traits can be facilitated by using molecular approaches, such as the construction of a high-density genetic map based on functionally anchored fiber genes and by subsequent molecular tagging of fiber QTLs for MAS. As a step toward this goal, we developed 1019 EST-derived, PCR-based, microsatellite markers with a high interspecific transferability from ~14,000 consensus sequences based on fiber ESTs from *G. arboreum*, and a genetic linkage map composed of 121 new fiber EST-derived and 88 public genomic microsatellite markers was constructed using an interspecific (*G. hirsutum* × *G. barbadense*) RIL population.

Data mining of fiber EST consensus sequences containing microsatellites revealed that approximately one-half (48.5%) of the microsatellites identified were trinucleotide repeats. The preponderance of trinucleotide motifs in EST-derived SSRs has been noted in humans (Subramanian et al. 2003) and plants (Chin 1996; Morgante et al. 2002; Temnykh et al. 2000; Thiel et al. 2003), including *Arabidopsis* (Morgante et al. 2002). The frequency of trimeric SSRs in coding regions is attributed to the hypothesis that selection against frameshift mutations suppresses the expansion of non-trimeric nucleotide repeats (Metzgar et al. 2000). Our observation that AT and AAG occur as the most abundant SSR motifs in cotton fiber ESTs is concordant with the generalized distribution found in plant genomes (Gupta et al. 1996; Eujayl et al. 2002). AT is also a typical dimeric repeat in non-transcribed regions of *Arabidopsis* (Morgante et al. 2002). In contrast, the AT motif is under-represented among SSRs derived by EST mining in barley (Thiel et al. 2003), rice (Temnykh et al. 2000), maize (Chin 1996) and *Arabidopsis* (Cardle et al. 2000). The representation of the trimeric repeat AAG in cotton differs from that in the other plant species described above, with the exception of *Arabidopsis* (Cardle et al. 2000). CCG was the most frequent trimeric repeat in barley (Thiel et al. 2003) and rice (Temnykh et al. 2000), and CGC in maize (Chin 1996). The differences in the type of abundant motifs may imply that the relative degree of representation by specific SSR motifs may vary across genomic regions (coding or non-coding) and/or different organisms. Unlike SSRs, allelic variation of CSRs has not been extensively evaluated in cotton and, although lower in overall frequency in the genome, should, in theory, provide a better source for detecting polymorphism. However, the polymorphism level resulting from the two marker types in this study were 46% for CSRs and 55% for SSRs. This difference could be attributed to the fact that the repeat length cut-off values applied to the two markers were also different (> 10 vs. > 12 bp). In addition, simple SSR motifs are more likely to undergo slippage during transcription than CSRs are. Allelic variation at CSR loci may result from In/Dels or single base-pair mutations in the repeats, or the sequence immediately flanking the repeats, in conjunction with differences in the number of repeat units (Urquhart et al. 1994; Garza and Freimer 1996; Brinkmann et al. 1998a, b; Lin et al. 1998; Bull et al. 1999). Although PCR fragment length variation is not detected among CSR alleles, sequencing could reveal allelic differences between PCR amplicons that are identical in length.

The high interspecific transferability of microsatellite markers was demonstrated by successful PCR amplification across six different species of *Gossypium* with 82.7% of the primer sets tested. Higher cross transferability of EST-derived markers compared to genomic microsatellites has also been demonstrated in barley (Thiel et al. 2003) and sugarcane (Cordeiro et al. 2001), and attributed to sequence conservation of coding regions of the genome. The microsatellites reported here may be useful for introgressing alleles for fiber traits of interest from diploid species and exotic germplasm collections into modern cotton cultivars through MAS, and may contribute to understanding of the evolution of allotetraploid species following their divergence and domestication. On the other hand, the conserved nature of EST-derived microsatellites reduces the frequency of polymorphism. Only 19.8% of the (1019) EST-derived microsatellites analyzed in this study were polymorphic between the two parents of the interspecific mapping population, *G. hirsutum* and *G. barbadense*. A similar level of polymorphism (18.2%) was reported by Han et al. (2004) using fiber EST-SSRs, while SSR-based markers from an SSR-enriched genomic DNA library reported 49 and 56% polymorphism, respectively, between the two species (Reddy et al. 2001; Nguyen et al. 2004). One reason for the relatively low polymorphism rate of EST-derived microsatellites could be natural selection against frameshift mutations that limit

microsatellite expansion in coding regions (Metzgar et al. 2000). Polymorphism in this study may be underestimated because of the possibility that some of the SSRs classified as monomorphic might actually be polymorphic if separated on a high-resolution (1 bp) system such as polyacrylamide sequencing gels. In this study, electrophoresis on 3% SFR agarose gels reliably detected allelic differences >5 bp in length, especially for PCR amplicons ranging in size between 250 and 300 bp. In addition, digestion of PCR products with diverse restriction enzymes may reveal DNA sequence polymorphisms within restriction sites. This cleaved amplified polymorphic site (CAPS) marker technique may be an alternative way to conveniently convert monomorphic SSR markers for important functional genes into polymorphic markers (Chee et al. 2004). Single-strand conformation polymorphism analysis can also reveal single nucleotide polymorphisms without the need to sequence monomorphic DNA fragments (Lucas et al. 1997).

The genetic map reported in this study is the first fiber EST-microsatellite map constructed using a RIL population in cotton. A RIL mapping population provides several advantages over F_2 or BC_1 populations derived from a single meiotic event. Once segregation is completed or nearly completed in RILs, the population can be permanently used for diverse mapping studies for different traits. According to Haldane and Waddington (1931), the proportion of recombination between two linked loci in an inbred population is twice that estimated using a single meiotic event. This theoretical relationship suggests that recombination between closely linked markers is more readily detected by RILs (Burr et al. 1988). The high percentage of loci with significant segregation distortion observed in this study implies that a genetic mechanism may exist for favored transmission of particular genomic segments (Lu et al. 2002). Some of these loci showed allelic frequencies skewed towards the maternal (TM1) parent (possibly due to pollen lethality caused by deleterious genes in the male parent 3-79), and were partitioned into three substantial linkage blocks. Similar observations have frequently been reported for interspecific crosses in cotton and other crops (Nikaido et al. 1999; Jiang et al. 2000; Ky et al. 2000; Rong et al. 2004; Shen et al. 2005). Cytological assignment of microsatellite loci using aneuploid stocks provides a valuable means for associating genetic linkage groups to specific chromosomes. Our genetic map located 108 genes expressed during fiber morphogenesis (Arpat et al. 2004) to 13 chromosomes of the A-subgenome (At), and ten chromosomes of the D-subgenome (Dt) (Fig. 3). Although D-genome diploid species and ancestors of the Dt-genome in allotetraploids do not produce spinnable fibers, many fiber EST loci were mapped on the Dt-genome in this study and in that performed by Han et al. (2004). Indeed, many QTLs that positively affect fiber quality have been detected in the Dt genome (Chr. 14, 15, 18, 20, 22, 23, 25, 16, and 26; Kohel et al. 2001; Paterson et al. 2003; Shen et al. 2005; Ulloa et al. 2005b), a fact that underscores the contribution of the Dt genome to cotton fiber development and agronomic properties in cultivated tetraploid species.

To date, few cotton genetic maps include QTLs for fiber quality (Kohel et al. 2001; Paterson et al. 2003; Shen et al. 2005; Ulloa et al. 2005b). Fiber EST-derived markers are especially informative, since the marker polymorphisms can cause changes in gene function, and thereby directly influence phenotypic variation, as evidenced for amylase content controlled by the *Waxy* gene in rice (Ayers et al. 1997). In our preliminary QTL analysis, only one QTL was identified for 50% *Lf* on Chr. 02. The paucity of QTLs declared at a significant permutation threshold may be due to the relatively small genome coverage (approximately 27%) of the map used for QTL analysis, and/or coupled to phenotypic data obtained from the F_2 progeny instead of RILs. Nevertheless, chromosomal locations of the fiber QTLs reported in the previous studies (Kohel et al. 2001; Paterson et al. 2003; Ulloa et al. 2005) strongly suggest that the other eight QTLs detected at LOD≥2.0 may represent true QTLs. As mentioned earlier, QTL analysis based on phenotypic data, and RFLP and RAPD markers, from F_2 progeny used to develop the RILs in this study, reported four QTLs for *Sf* on Chr. 03, 14, 15 and 25, and six QTLs for *Ff* on Chr. 01, 02, 03, 12, 16, and LG.D01 (Kohel et al. 2001). Using the same F_2 phenotypic data, but different marker genotypes from RILs in this study, putative QTLs for *Sf* (LOD = 2.7) and *Ff* (LOD = 2.7) were also identified on Chr. 15 and 03, respectively. In addition, Ulloa et al. (2005b) reported an important fiber QTL for *Ff* on Chr. 03. QTLs for *Ef* (LOD = 2.1) and *Sf* (LOD = 2.6) detected on Chr. 15 and 18, respectively, in this study, were also detected on the same chromosomes by Paterson et al. (2003) using an independent interspecific F_2 population. The location of fiber QTLs to the same chromosomes in different mapping populations supports the QTL assignments in this study, despite the low statistical support from permutation tests. Further validation of these QTLs awaits phenotypic data from RILs, which is expected to become available from the 2005 growing season at two locations (in California and Texas).

Finally, the comparison of EST sequences and map positions with NAU SSRs proved that the majority of MUSS/MUCS markers were developed from different ESTs and the homologous markers were in agreement for genetic map locations. The reasons for the low level of overlap between the EST-derived SSR maps include the use of different data sets (ESTs versus assembled consensus sequences) developed using different programs, and searched for non-overlapping target motifs. While ~40% of the ESTs selected by Han et al. (2004) carried hexanucleotide motifs not included in our study, the work reported here also included CSRs. The two studies also used different cut-off values for primer design—repeat length >18 bp in Han et al. (2004) and >10 bp in this study. Twenty out of 29 MUSS/MUCS loci homologous to cDNA-RFLPs mapped by Rong

et al. (2004) were not mapped at the same or homoeologous chromosomes in this study. The differences in alleles detected by cDNA-RFLP and EST microsatellites (MUSS, MUCS, and NAU) can be explained by the nature of the marker types employed for each map. It is possible that cDNA-RFLP loci not detected by EST-SSRs may be duplicated sequences that were monomorphic or not amplified from our PCR-based markers, or the detection of the loci was due to non-specific hybridization between the probe and genomic DNA with sequence similarity.

Acknowledgements This research was supported, in part, by Cotton Incorporated Project Nos. 03-398 and 03-321CA (to M.U.) and the NSF Plant Genome Research Program (DBI 9872630 to T.A.W). We would like to thank James Frelichowski, Ravinder Gill and Gregory Allen for running agarose gels. We would like to gratefully acknowledge Cotton Inc. for its support on this project. We also would like to acknowledge the U.S.-Egypt Joint Cotton Project (BIO7-001-015) for providing partial support of Magdy S. Alabady at the USDA-ARS, WICS, Cotton Enhancement Program.

References

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage map. Genome 36:181–186

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA (2004) Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol 54:911–929

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Ayers NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellite and single nucleotide polymorphism differentiate apparent amylase classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94:773–781

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 32:314–331

Brinkmann B, Junge A, Meyer E, Wiegand P (1998a) Population genetic diversity in relation to microsatellite heterogeneity. Hum Mutat 11:135–144

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998b) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Gene 62:1408–1415

Bull LN, Pabon-Pena CR, Freimer NB (1999) Compound microsatellite repeats: practical and theoretical features. Genome Res 9:830–838

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Chee PW, Rong J, Williams-Coplin D, Schulze SR, Paterson AH (2004) EST derived PCR-based markers for functional gene homologues in cotton. Genome 47:449–462

Chin ECL (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 156:847–854

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative traits mapping. Genetics 138:963–971

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *Erianthus* and *Sorghum*. Plant Sci 160:1115–1123

Eujayl I, Sorrells ME, Wolters P, Baum M, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Res 6:211–217

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 70:45–53

Haldane JB, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374

Han Z, Guo W, Song X, Zhang T (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Genet Genomics 272:308–327

Ji SJ, Lu YC, Feng JX, Wei G, Li J, Shi YH, Fu Q, Liu D, Luo JC, Zhu YX (2003) Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucleic Acids Res 31:2534–2543

Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor Appl Genet 100:409–418

Kohel RJ, Yu J, Park Y-H, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163–172

Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, Charrier A, Hamon S, Noirot M (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee. Theor Appl Genet 101:669–676

Lacape J-M, Nguyen T-B, Thibivilliers S, Bojinov B, Courtois B. Cantrell RG (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* × *Gossypium barbadense* backcross population. Genome 46:612–626

Lin L, Jin L, Lin X, Voros A, Underhill P, Mignot E (1998) Microsatellite single nucleotide polymorphisms in the HLA-DQ region. Tissue Antigens 52:9–18

Liu S, Saha S, Stelly D, Burr B, Cantrell RG (2000) Chromosomal assignment of microsatellite loci in cotton. J Hered 91:326–332

Lu H, Romero-Severson J, Bernardo R (2002) Chromosomal regions associated with segregation distortion in Maize. Theor Appl Gent 105:622–628

Lucas M, Munoz C, Pintado E, Solano F (1997) Highly informative single-stranded conformation polymorphism (SSCP) of short tendem repeats in DNA identification. J Forensic Sci 42:118–120

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutation limits macrosatellite expansion in coding DNA. Genome Res 10:72–80

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet 30:194–200

Nikaido A, Yoshimaru H, Tsumura Y, Suyama Y, Murai M, Nagasaka K (1999) Segregation distortion for AFLP markers in *Cryptomeria japonica*. Genes Genet Syst 74:55–59

Nguyen TB, Giband M, Brottier P, Risterucci A-M, Lacape J-M (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor Appl Genet 109:167–175

Paterson AH, Saranga Y, Menz M, Jiang C, Wright RJ (2003) QTL analysis of genotype × environmental interactions affecting cotton fiber quality. Theor Appl Genet 106:384–396

Percival AE, Wendel JF, Stewart JM (1999) Taxonomy and germplasm resources. In: Smith CW, Cothren JT (eds) Cotton: origin, history, technology, and production. Wiley, New York, pp 33–63

Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: a new class of genetic markers in cotton. J Cotton Sci 8:112–123

Reddy OUK, Pepper AE, Abdurakhmonov I, Saha S, Jenkins JN, Brooks T, Bolek Y, El-Zik KM (2001) New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. J Cotton Sci 5:103–113

Reinisch A, Dong J-M, Brubaker C, Stelly D, Wendel J, Peterson A (1994) A detailed RFLP map of cotton (*Gossypium hirsutum* × *G. barbadense*): chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Rong J, et al (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy OUK, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grapes EST. Theor Appl Genet 100:723–726

Shen X, Guo W, Zhu X, Yuan Y, Yu JZ, Kohel RJ, Zhang T (2005) Molecular mapping of QTLs for qualities in three diverse lines in Upland cotton using SSR markers. 15:169–181

Stam P, Van Ooijen JW (1995) JoinMap™ Version 2.0: software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen

Stelly DM (1993) Interfacing cytogenetics with the cotton genome mapping effort. In: Herber DJ, Richter DA (eds) Beltwide cotton conference, New Orleans, January 10–14, 1993. National Cotton Council of America, Memphis, pp 1545–1550

Subramanian S, Madgula VM, George R, Kumar S, Pandit MW, Singh L (2003) SSRD: simple sequence repeats database of the human genome. Comp Funct Genomics 4:342–345

Tautz D, Schotterer C (1994) Simple sequences. Curr Opin Genet 4:832–837

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Ulloa M, Meredith WR Jr (2000) Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. J Cotton Sci 4:161–170

Ulloa M, Meredith WR Jr, Shapply ZW, Kahler AL (2002) RFLP genetic linkage maps from F2.3 populations and a joinmap of *Gossypium hirsutum* L. Theor Appl Genet 104:200–208

Ulloa M, Mc Stewart J, Garcia CE, Godoy AS, Gaytan MA, Acosta NS (2005a) Cotton genetic resources in the western states of Mexico: in situ conservation status and germplasm collection for ex situ preservation. Genet Resour Crop Evol, in press

Ulloa M, Saha S, Jenkins JN, Meredith WR, McCarty JC, Stelly MD (2005b) Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) joinmap. J Hered 96:132–144

Urquhart A, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. Int J Leg Med 107:13–20

Van Ooijen JW, Maliepaard C (1996) MapQTL™ Version 3.0, sotfware for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, The Netherlands

Wang S, Basten CJ, Zeng ZB (2001–2004) Windows QTL Cartographer 2.0. Department of Statistics, North Carolina State University, Raleigh, NC

Weber JL (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics 7:524–530

Wendel JF, Cronn RC (2003) Polyploidy and evolutionary history of cotton. Adv Agron 78:139–186

Wilkins TA, Arpat AB (2005) The cotton fiber transcriptome. Physiol Plant, in press

Wilkins TA, Arpat AB, Sickler BA (2005) Cotton fiber genomics: developmental mechanisms. Pflanzenschutz-Nachrichten 58:119–139